DOCUMENT RESUME

ED 037 386 SP 003 601

TITLE	Untangling the Tangled Web of Education. Research and Measurement Considerations Related to Assessing Children's Development in Interaction with School, Family, and Community Influences.
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	69
NOTE	41p.; A special symposium sponsored by the National Council on Measurement in Education, New York, November 1, 1968
AVAILABLE FROM	Educational Testing Service, Princeton, N.J. (free)
EDRS PRICE	EDRS Price MF-$0.25 HC-$2.15
DESCRIPTORS	Community Role, *Disadvantaged Youth, *Early Childhood Education, *Educational Research, Family Role, Human Development, Longitudinal Studies, Measurement, Research Design

ABSTRACT

This booklet contains the seven papers presented at the symposium: (1) "Early Schooling: What Is It All About?" by Marshall P. Smith, Trenton State College, who approaches the subject as "one of the preeminent tools for human survival"; (2) "The Family and Community: What Are Their Roles in the Educational Process?" by Melvin Tumin, Princeton University; (3) "The Child: His Cognitive, Personal-Social and Physical Development--A False Trichotomy?" a discussion of the need for integration of the three in the educational process, by Edmund W. Gordon, Columbia University; (4) "How Are Measurement Strategies Related to Models of Human Development?" in which Walter Emmerich of Educational Testing Service presents candidate models of human development, all calling for longitudinal research; (5) "Can You Do Real Research in the Real World?" a discussion of generalizability and interpretability in choosing research strategies, by Samuel Messick, Educational Testing Service; (6) "The ETS-OEO Longitudinal Study of Disadvantaged Children," a presentation of aims and design for the planned 6-year study of children from age 3 to grade 3, by Scarvia B. Anderson, Educational Testing Service; and (7) "The Scientific and Social Significance of the Longitudinal Study of Disadvantaged Children" by John W. McDavid, University of Miami. (JS)

untangling the tangled web of education

Research and measurement considerations related to assessing children's development in interaction with school, family, and community influences

A special symposium sponsored by the National Council on Measurement in Education

EDUCATIONAL TESTING SERVICE ETS PRINCETON, NEW JERSEY

An Explanatory Note

Toward the end of October or early in November each year, several educational and professional organizations hold conferences in New York City for an exchange of information about matters pertaining to educational research and measurement. On November 1, 1968, some 200 of these educators and psychologists attended a special symposium sponsored by the National Council on Measurement in Education, in conjunction with the conferences of the Educational Records Bureau and Educational Testing Service.

Perhaps they were lured by the symposium's title – "Untangling the Tangled Web of Education" – and its distinguished speakers. Or perhaps they had heard that Educational Testing Service, under a grant from the Office of Economic Opportunity, was about to embark on a six-year longitudinal study of disadvantaged children and their first school experiences. In any case, they came, they listened and they asked questions. They also indicated that they hoped the symposium papers would be made available to everyone interested in problems of educational research and evaluation in today's "real world."

This booklet presents the papers as they were delivered that day last fall. Scarvia Anderson of Educational Testing Service and Jerome Doppelt of The Psychological Corporation, chairmen of the informal symposium, planned the program. Mr. Doppelt introduced each speaker (see Contents page) and kept the proceedings right on schedule.

The last speaker was John W. McDavid, former Director of Research and Evaluation for Head Start in the Office of Economic Opportunity. Mr. McDavid related some of the theoretical and practical issues that had characterized the earliest discussions of the design and objectives of the ETS-OEO Longitudinal Study. He described the study as "action research" in which research and evaluation would be combined. He frankly stated that "we do not expect execution of the longitudinal study to be without problems" – but he also characterized it as "potentially the most significant single piece of educational research undertaken in this decade."

April 1969

Contents

Early Schooling: What Is It All About?

Marshall P. Smith, Trenton State College

I am asked to face the question "Early schooling – what is it all about?". There has been so much said, and so often, about early schooling that in trying to add anything I feel somewhat at a loss.

When I was a very small boy we had a kaleidoscope in the family – you may remember those gadgets. Looking through an aperture down a cylinder about the size of an oatmeal box you could see, when you turned the cylinder, constantly changing and ever-new patterns of colors in symmetrical arrays. It was a gratifying activity. I felt truly creative, the producer of uniqueness in structure – that is, until a cynical older sister pointed out that each new design was done with precisely the same colored beads and that the new symmetries were all done with mirrors.

In trying to say something on early schooling I seem to be playing with a kaleidoscope. I'm supposed to say, "Look at this great new analysis!" But I know and you know it is pretty much going to be the same old beads simply reflected differently in the same old mirrors. Originality is hard to come by.

What then should I say in this brief time that might have some significance? I shall skip the promise of valuable community involvement the program offers. I shall avoid the argument that early schooling relieves the problems of working mothers.

I'm going to skip, since you've heard it all already, the arguments about preparation for formal schooling so that the children's later school experience will be rewarding rather than frustrating.

I shall skip, too, the argument that early schooling will help spot potentially superior students before they get caught in the massive maw of the traditional system.

Instead I shall make it my major point that when we deal with early schooling we are dealing with one of the preeminent tools for human survival.

Children in our society customarily enter the first grade at a mean age of about 6 years and 3 months. Early schooling, which is my topic, has come to mean the two years before first grade, including kindergarten and the year prior to kindergarten, the entering age being about 51 months – that is, 4 years, 3 months.

When you think of that age it will strike you as very young – but it is no younger than the age for children entering suburban nursery schools. Plato, modeling his early childhood education after Sparta,

placed the start of supervised schooling at age 4. Small Navajos got very small but functioning bows and arrows around this same early age, often from their grandfathers, and I imagine there are some of you who at four trailed behind mother to help in the kitchen garden or behind father to help feed the cows. Many of you, I imagine, cannot even remember how early you started performing in miniature the activities of the important adults about you, activities that generated feelings of autonomy, initiative, and the beginning sense of competence.

The church has long recognized the necessity for "getting them young" if desired basic character traits are to be firmly rooted, and we know how successful it was for centuries. And who was it said the hand that rocks the cradle is the hand that rules the world? The U.S.S.R. started early, and has continued, state-sponsored day nurseries for working mothers, and we have read how strong in these is the emphasis on developing those character traits and values that would characterize the "good" soviet citizen. The kibbutz in Israel is avowedly in the business of molding those attitudes toward the self and the society that will support and foster the development of the community. So "early schooling," either formal or informal, is nothing unique. In fact it has been so widespread that we must assume it has a valid function.

I argue that early schooling of some sort is needed by every society if it is to start a new generation on the road to mature productivity. Each new generation needs this early experience if its members are to mature in such a way as to preserve and advance the values of the culture and to become competent in performance of the tasks it will be called upon to perform. But the early *institutionalized* school – age 4 to age 6 – has only recently become a social necessity. Where the essential developmental tasks were built into the social structure, institutionalization of early schooling was unnecessary. The small boy who helped in his father's smithy or cobbler shop, or helped with the chickens and the pigs, was successfully mastering – if his father was wise – the critical personal tasks of achieving autonomy, initiative, and a sense of competence. At the same time he was learning the foundations for attitudes of workmanship and productiveness, and thus citizenship. Just so were the Navajo and the Spartan; just so is the child of the kibbutz.

We speak often and with fervor in education of the need to focus upon individual needs and the need to foster the highest development of the individual. In a social evolutionary sense these great values are – I hate to say this – incidental, or, to put it more softly, a desired but accidental dividend of the much more important major aim of

education. *That major aim is meeting society's need for competent and self-respecting citizens.*

The problem here, of course, is that to a large extent the modern society of the city offers to the young no valid equivalent to the family learned self-perceptions and identifications of earlier generations. In the absence of the semiautomatic early schooling of the nuclear family or home shop or tribe we find developing in our cities, and perhaps our suburbs, a peer-oriented, other-directed street culture. This culture is not only irrelevant to the needs of the society but is often positively inimical both to the general welfare and the welfare of the individual. The home, often, can no longer do the necessary job especially for the very young, and society is forced, therefore, for its own preservation to invent a way of meeting the crisis of the irrelevancy of early experience.

This problem of the irrelevancy of modern early experience to society's needs is compounded with changes in technological skill requirements such that what little *is* modeled for the very young becomes obsolete before those skills can be meaningfully utilized by the growing child and the young adult.

Frequently the problem is further aggravated in the case of boys by the absence of valuable male identification figures; or indeed by the *presence* of identification figures of negative value. It is hard to develop the rudimentary feelings of competence and worthiness if there are few models to become attached to. It is often the case that the value of males is decried by important figures in the home, with the boy growing unconsciously defensive and hostile as he comes to recognize his own devalued maleness.

The problem is further compounded, in the case of black children, by a white racism that simply and matter-of-factly takes it for granted that the black child is really not worth very much and conveys this message through every medium. Don't think 4 or 5 years of age is too early to learn this message of unworthiness. Any of you who know an undervalued child can read the signs. Imagine a whole culture taking this callous approach. Note that the white racism does its work for the most part without real emotional rejection. It simply takes the black child as naturally worth not very much. The massive effect of this cold approach on the child is to define his one and only world to be this way – a cold world that cannot really be fought, or a world where fighting back generates only further defeat and guilt.

If some of these are characteristic experiences for all small children, and if *all* are characteristic experiences for some, the conclusion is compelling. Society must devise and apply massively a

program of universal early schooling that will reach all the children in these critical early years, structuring a learning environment that will be truly relevant to future maturity.

If society is to serve its own need to develop productive citizens with competence and self-confidence, and earn, as a dividend, citizenry who value themselves and others, the most critical time is the time of early schooling. This is the period when the developmental tasks of establishing autonomy and initiative are faced. This is the period that becomes the base, in turn, for the achievement of industry and a valid sense of identity. In the absence of a mastery of the critical early tasks, the developmental alternatives are self-doubt, guilt, feelings of inferiority and desperate anger.

This is what early schooling is all about. Our society must serve itself in the future through educating children now When I said earlier that early schooling was a social necessity, I meant it. It is one of our few outs if our country is not to meet catastrophe. If I could decide, I would not start early schooling at 4 years of age, I would go right back to 3 years or even 2. My motherly secretary says to this: "Why, they are only babies!" To which I reply: "Amen."

The Family and Community: What Are Their Roles in the Educational Process?

Melvin Tumin, Princeton University

It is sociologically axiomatic that when a number of parties are involved in any social enterprise, and when the enterprise fails, each party will lay maximum blame for the failure on the others, and will assume only minimum blame, if any, for itself. As a corollary, it follows that the *official* verdict of guilt for failure will be imposed on that party who is weakest or least able to fend off the imposition of the official stigma.

A variety of circumstances have joined today to produce the widespread notion that the American public schools have failed, especially with regard to children of lower socioeconomic families, and most especially in the case of Negro children. This is a comparatively new development in educational jurisprudence. For up until recently, the schools as such were not judged to be failing. Rather, blame was officially imposed on those children who did not manage, for one reason or another, to live up to official standards and expectations. Being powerless, relative to all other parties, children have had no alternative but to accept and suffer the ruling, embodied in their report cards, and ritually celebrated in honors assemblies and boasting matches at conventions of principals and superintendents. At these tribal gatherings the managers of schools or school systems deftly lay claim to their entitled places on the pecking order of school prestige, supporting their claims by reports of the numbers of their students who score above national averages, or who win prestigious national advertising campaigns, called Merit competitions, or who were admitted to "high ranking" private schools.

Comparable public celebrations of the success of some and the failure of many children have been built into the very heart of school operations, in the form of promotion and retention at the end of term or year (or the equally monstrous procedure of automatic promotion for all); in the assignment of honors students to the so-called best teachers, as a reward for teacher excellence; in the tracking and grouping of children into so-called ability groups; in the sharp distinctions between the college preparatory curriculum and all other curricula in the secondary schools. There are numerous other evidences of the deep commitment of American education to blaming children for failing to learn as much as the "standards" demand that

they shall. But these will suffice to indicate the depth and ubiquity of that commitment.

But all of this seems very much in the process of change. For now, various segments of the public, alerted to the dismal regularity and predictability of the "failure" of large numbers of children, have taken turns laying the blame at each other's doorstep. This has had the benign effect of providing some of the children, for the moment at least, with a respite from daily involvement in failure, shame, and public degradation.

Thus, for nearly 20 years, starting just after World War II, the teachers of America, and *their* teachers, were attacked from all sides for the educational failures of children. Then, for a brief moment, until a temporarily successful counterattack was launched, the families of the children, especially of black children, were held to be essentially defective – through no fault of their own, but defective, nevertheless, in educationally crucial regards.

Most recently, it is a combination of the educational establishment (whatever many things that means) and of the corollary lack of community control of the schools that has been made the major scapegoat. In the judgment of some spokesmen in the black community, there has been a conspiracy, sometimes averred to be deliberate, to keep black children uneducated; or worse still, to "murder" them, at least educationally. The particular rhetoric or claimed level of injury does not really matter. What is more significant, here at least, is the fact that *not* the children and *not* the families, and *not* their *own* communities, but rather the absence of community involvement and rule are held to be responsible for the children's failing to learn. By explicit implication, the assumption of community rule of schools by indigenous members of the community is militantly claimed to be a *sine qua non* for decent education.

One can hardly expect participants in emotionally-charged political struggles to be majestically objective. It is no surprise, therefore, that the respective parties to these disputes should seem unable to agree on what are the relative weights of importance of family, community, and school in the determination of the educational outcome of the children, for whom the schools presumably were intended in the first place. And since almost no one has asked for clear-cut operational specifications of what "failure" means, or whether it is a worthy educational concept at all, the task of adjudicating the disputes has been made even more difficult.

How shall we know, five years from now, whether experiments in community control of schools have been successful, if we don't agree at the outset as to what we mean by success and failure? How much

is it reasonable to expect which children to learn, in what kinds of schools, with what kinds of curricula, and as measured by which instruments? Until we can get some moderately agreeable answers to those questions, we can not expect to introduce much rational order into current educational debates.

Whatever our supreme ignorance on many key educational questions may be, it seems quite clear, to the majority of the educational research community at least, that family life, community organization, and the schools themselves are all contributors to the educational outcomes of the children. It is certain that we do not know in what relative proportions these three major bundles of factors contribute their influences. But we *do* know, from numerous researches, that differences in the school performance of children are variably attributable to differences in factors located within these three domains.

We know, for example, from Benjamin Bloom's assessment in *Stability and Change in Human Characteristics*, that school-related abilities are nourished and shaped decisively in early childhood. So, presumably the family *is* a decisive variable, all protests to the contrary notwithstanding. We know, too, from such researches as the Coleman report on *Equality of Educational Opportunity*, that, at least as measured by available instruments, the quality of teachers has some influence, as does even more the socioeconomic level of the children with whom one attends school. Presumably, then, the culture of the peer group is one that has its own norms, that conveys its own standards of aspiration and achievement, and that acts as a community of educational support – or its opposite. But from the same report we also know that the socioeconomic standing of the family, and all that it implies, is more important than any other single set of factors in shaping the educational fate of the child.

Since communities tend to be relatively homogeneous in their socioeconomic composition, it stands to reason that lodged in the character of the community are some crucial factors relevant to the outcome of the child in the schools. Among the most important of these may be the self-assurance and knowledgeability of the parents of the community in the management of the educational careers of their children, including the ability to forcefully impose their notions of school conduct on the professionals who conduct the day-by-day operations of the schools.

We know, too, from such penetrating researches as those of Robert Rosenthal that the preconceptions of teachers as to how students ought to perform in the schools have extraordinary influences upon the actual outcomes. Here, then, such factors as teacher facilitation

of promising children; systematic ignoring or downgrading of the performances of so-called unpromising children; and the reciprocal energizing of the favored and demotivation of the unfavored, all contribute, probably, to the operation of this process of self-confirming hypotheses.

It is no accident, therefore, that any attempt to evaluate school or preschool programs ought to be concerned with what their effects may be on a wide variety of children's behaviors and performances, and even more importantly, with what it is about these programs that makes the difference, if any, in their educational effects. It is logical, therefore, that we should examine in considerable detail a range of possibly important factors connected with the home life, the community life, and the actual educational or school experiences of the children exposed to these programs.

We should be asking: What kinds of families and communities do the children come from? Are there educationally relevant supports, such as adequate light, privacy, nourishment? Can the parents play the role of auxiliary teachers, as so many educated parents can and do? Are there living models in the homes and the daily lives of the children of the relevance of educational striving, of diligence and regularity, and of systematic attention to tasks? Is there incidental yet regular and nourishing interaction among siblings, and between them and parents, that contributes to the intellectual and emotional health and development of the child? Are the educational artifacts, such as books, encyclopedias, magazines, present to any significant degree? Is there, in short, any real continuity between the ambience of the school and that of the home?

In probing all of these questions, of course, we shall find variations among families, and we will want to know whether the measured differences in cognitive and affective development of the children over the years can be attributed to various combinations of factors indicated above.

Since, too, it has been so forcefully insisted in numerous quarters that community control or participation in the management of schools is crucial in various ways, we should look intensively at the structure and functioning of the community in which the child and his family reside. Do the parents have a sense of their effective community? Are they aware of the facilities or the lack of them in their neighborhoods? Do they know to whom to turn for help in a number of contexts where they need help? Do they have a sense of their capacity to shape the community resources to meet their needs?

Where families differ in these regards, as they surely will, even within the same communities, we shall want to know what bearing

these differences have, if any, on the self-concept of the parent and, derivatively, on the self-concept of the children. Do children whose parents feel more powerful and autonomous than others also feel more powerful and autonomous than children from families whose level of uncertainty and insecurity regarding the management of their affairs may be much higher? Are there differences in the school behavior and performances of children who come from areas where there are variable degrees of community organization and senses of competence? Where parents are more actively involved in school affairs, through visits, or organizational affiliations, or actual programs of school-community interaction, do their children show any significant differences in their own sense of their abilities and in their own conduct of their educational careers?

These are not easy questions. They are not easy to formulate; they are not easy to put into a form capable of being analyzed carefully and with some degree of precision; they are especially difficult in regard to assessing the respective influences of each of the numerous factors that will surely prove to be contingent upon each other. We should try to discover profiles and batteries of such contingent factors; we should try to understand the sequences in which they operate and then are operated upon; we should, in short, attempt to discover what it is about the family and community lives of children that may contribute to their cognitive and affective developments.

Needless to say, we are not likely to be surprised by the findings, whatever they may be. It could well be that energetic, self-confident, and active families may produce children who do not do significantly better in measured school tasks than families with considerably less of the apparently relevant characteristics. It may prove to be the case that there is very little, or to the contrary, very great relevance in the fact of absenteeism of fathers. Whether this absenteeism is structural or functional also may or may not prove relevant – structural through divorce or death or separation, or functional in the form of excessive dedication to work and marketplace, or in the form of educational incompetence. It may also turn out that an educationally nourished and nourishing family is of little or no avail if the child is exposed to an educationally depriving program and regimen in the school, just as it would be no surprise if the combination of supportive family life and energizing school program proved to be the most felicitous of all for the children concerned.

We stress these numerous possibilities of discovery mostly to indicate that the existing research is most indecisive indeed in what it tells us about these matters at the moment. In the same context, however, it is also crucial that everyone concerned should realize the

extraordinary importance of our being able to find out whether educational programs make a difference, and if so, how? How, in what they do, and how, in the context in which they operate: the context of the culture, structure and functioning of the school itself, the families, and the communities in which the schools and children are located.

The Child: His Cognitive, Personal-Social, and Physical Development – A False Trichotomy?

Edmund W. Gordon, Columbia University

With the vast increase in concern, effort, and money that has come to be focused on education in the past decade has also come an increased concern with evaluating the products of that effort. The National Defense Education Act was directed primarily at enhancing educational development in our most able students, and where evaluation was attempted it consisted largely of head counting: How many students did we reach? To what extent did we increase the number of students or scholars in the target discipline?

Much the same can be said for the related efforts of the National Science Foundation. However, as the nation turned its attention to expanding educational opportunities for less able students – more correctly, for students who do not manifest their potential in ways the school is accustomed to recognizing or, more colloquially, the socially disadvantaged – evaluative research problems took on a different character. Counting the number of persons served was relatively unimportant. Determining the impact of service or performance became an important issue. With the initiation of a variety of massive educational enrichment programs, the establishment of the anti-poverty program, including Head Start, and the implementation of the Elementary and Secondary Education Act, evaluative research in education experienced phenomenal growth. Overnight we fielded a new army of alleged experts in educational evaluation. (Your speaker, incidentally, marched in front with several other newly commissioned generals.) These investigators set about to crudely document the rapidly emerging programs and their impacts on children and youth.

The principal focus of this evaluative research was placed on changes in cognitive development as reflected in scores on standardized tests of intelligence and academic achievement. A review of many of the reports emanating from these studies reveals negligible gains as reflected by these criteria, but almost always a subjectively determined greater gain in emotional-social development and stability.

The narrowness of the output measures, typical of these first efforts, reflects a bias that has plagued educational evaluation. Although the goals of education tend to be stated in broad terms, when we come to assess education it is always to cognitive development and academic achievement that we first look for evidence of change. Too often we either stop with those first results

or turn with less rigor to look at other areas either as a second thought or as a rationalization for our failure to find more impressive evidence in the cognitive domain.

There are a few people in education who strongly criticize this cognitive emphasis and remind us that education is equally concerned with affective development. They plead strongly for greater attention to the social-emotional aspects of development. But we are not as confident about our instrumentation in this area as we are about measurement in the intellective area, and efforts at evaluation with an affective emphasis reflect this. The measures in the emotional sphere do not easily lend themselves to quantification. The problems of reliability and particularly validity of measures are even more complicated than those in the measurement of intellectual function.

But even if the technology of affective assessment were better developed, the affective emphasis would be no more appropriate to educational evaluation than the dominant cognitive emphasis. The processes of development, education, and learning can best be understood in the context of an interactionist or transactionist approach to the understanding of any phenomenon. In any process there are several mechanisms or elements interacting together. Education involves a multitude of transactions between what is indigenous to the learner and that which is provided in the learning experience. The learner brings to the learning task a physical self, an intellective self, and an emotional self. These several aspects of self interact not only among themselves but also interact with the effective environment, and these respective interactions are reciprocal, dialectical, and relative (reciprocal in the sense that the interactions are two-way – each aspect of self reacts to the environment and also acts on the environment to change it and change the subsequent interaction; dialectical in the sense that an interaction in one self-system and environmental system influences reactions or interactions in other self-systems and environmental systems; relative in the sense that interactions are always a function of a particular set of conditions, and a specific interaction has to be understood in relation to these conditions). To study the physical self, the intellective self, or the emotional self in isolation, then, may be convenient at times, but it is never adequate to fully understand the status or nature of development.

The problem in evaluative research is not to determine if a treatment has made a difference, but to explain the nature of the interaction between specific aspects of the treatment and certain aspects of the treated. It is out of this understanding that intelligent decisions can be made with respect to repeating, expanding,

modifying, or curtailing treatment. Given the transactional nature of the educative process and the very complex patterns of interaction and interpenetration among the many aspects of individuals and environments involved, it is clear that to trichotomize the child for purposes of education or the evaluation of educational effort is to defeat the purposes of both.

We certainly have learned that evaluation efforts that focus on a single aspect of the child's development have proved unsatisfying and relatively unproductive. Many of these efforts fail to be sensitive to developmental changes that parents and teachers know to be present. But the modest positive gains of many of these studies are not simply a product of too narrow an evaluative focus. Many of the programs simply have not produced highly significant developmental gains. The failure or modest success of the programs may also be due to narrowness of program focus.

Learning proceeds through the utilization and modification of basic cognitive systems, basic affective systems, and specific skills and content mastery systems. Investigations by Zigler suggest that these systems are not equally malleable. According to Zigler, the cognitive system may be the least plastic while the affective system, as represented by attitude, motivation, involvement, and so on, may be more subject to modification through educational intervention and environmental manipulation. Yet it is the cognitive system that the school and particularly programs of compensatory education have sought to modify. Most of our more sophisticated educational interventions have focused on the development of frontal attacks on basic cognitive processes, while in those programs where affective processes have been the target more pedestrian innovations dominate. In almost none of these programs is there to be found a creative marriage between the two. It may be that the evaluative research findings continue to be modest primarily because neither input programs nor assessment programs have appropriately integrated the three systems.

There are glimmerings of movement in this direction in some of the emerging programs designed to serve disadvantaged young people. In addition to a deep concern with better understanding the relationships between school, community, and family influences and the developmental process in children, it is in the interest of strengthening and accelerating the integration of cognitive, personal-social, and physical development in the educational process that Educational Testing Service is undertaking a longitudinal study of a group of children from age 3 through their first experiences in formal education. This study will seek to integrate the several aspects

of program input and the personal-social, intellective, and physical aspects of development at the levels of process observation, qualitative assessment, and transactional analysis as well as at the level of interpretation. We recognize that evaluative research may not only provide some answers, but may also influence the direction of movement in the institutions and programs studied.

How Are Measurement Strategies Related to Models of Human Development?

Walter Emmerich, Educational Testing Service

As a developmental psychologist I am naturally interested in the ways that knowledge about human development contributes to a better understanding of the educational process. However, I plan to speak more broadly today about certain implications of the developmental point of view for conceptualization and measurement in educational research. My central theme is that systematic applications of developmental theory can increase the utility of our measures. In exploring this theme, I will be talking more about strategies of measurement than about its tactics and technology.

The developmental psychologist holds an image of the developing person as a system of interrelated functions that grow, differentiate, and become integrated and reorganized throughout the life span. Over the years this image has been translated primarily into cross-sectional studies in which specific functions and processes are compared across age periods. Recently, however, it has become increasingly apparent that certain developmental phenomena can be uncovered only through longitudinal designs in which repeated measures are taken at two, or preferably at several, age periods.

This trend is an inevitable one, for while all psychological measurement refers to discrete units of behavior assessed at some point in time, developmental constructs also refer to mechanisms and processes that link units of behavior over time in the same persons. The argument here is not primarily that longitudinal studies have value because they solve certain sampling problems, or because they increase the efficiency of statistical tests. Anyone who has engaged in longitudinal research will be the first to note that this approach raises more methodological problems than it can be expected to solve to our complete satisfaction. I am reminded here of the time when, after asking graduate students to discuss the pros and cons of the longitudinal method, I received a merciless barrage of reasons why this approach was methodologically unsound. Many of their reasons were correct, of course, but they were also largely irrelevant because they did not speak to the substantive developmental questions.

Fortunately, we are in a much better position today to judge the scientific gains that accrue from longitudinal studies, thanks primarily to such major efforts as those conducted at Berkeley and Fels. Indeed, we currently face the opposite risk of becoming oversold

before we know precisely what it is we are buying. I believe that we are now in a hazardous period during which we are tempted to move too rapidly from the general developmental image mentioned earlier, which is a beautiful image, to the technology of measurement. I believe that we still need to bridge the gap between our broad concepts of development and our impulse to measure, a step that is difficult and tortuous. What I am saying, then, is that while longitudinal designs are probably essential to provide a thorough test of any developmental theory, we still need to clarify which properties of these designs are relevant to which theories.

Let me illustrate how longitudinal designs can serve different functions for different models of development. Consider first how a univariate trait theorist might use longitudinal data. His reason for seeking repeated measures over time on the same trait might well be to determine the stability of the trait. He would argue, with some justification, that the very existence of the trait as a characteristic of human variability depends upon demonstrating the presence of reliable individual differences within several age periods and stability between age periods. Empirical demonstration of trait stability is important here because it relates to the theoretical quest to establish the trait's universality. The appropriate measurement strategy is to assess the same underlying characteristic at various age periods in the same persons, and then correlate between age periods. Identical instruments might be applied at all age periods, or these instruments might differ only with respect to difficulty level. In either case, the essential criteria for measurement are to tap the same content at each age level, and to end up with reliable individual differences at each age period.

Now consider the more complicated task of the multivariate trait theorist. Like his univariate brother, he is interested basically in establishing trait universality, but he goes one step further. He will argue for the greater theoretical power of the multivariate approach because the hypothesis of trait universality calls for generality across discrete attributes as well as across developmental periods. Here the measurement goal is to tap the same broad dimensions of individual difference at each age period under study. Recognizing that the behavioral manifestations of any general dimension will change as a function of age, this theorist will measure different behavioral indicators of the same dimension at each age period, with, of course, as much overlap as possible in the measures used at adjacent age periods. Through such a strategy, he is in a good position to test the hypothesis that a trait arises early in life and develops by a process of shedding early age-specific manifestations of the trait in exchange for

later age-specific manifestations of the same trait.

There are many variants of both the univariate and multivariate models just described, and I will briefly mention one of them. Suppose in the multivariate model it turns out that a specific attribute is found to belong to one general dimension at one age period but to another dimension at a later period. Suppose further that this state of affairs holds for a set of attributes that share a common meaning. Might this not be evidence for true dimensional change rather than universality throughout development? Here, then, is another possible model, not ordinarily considered by most trait theorists, but perhaps worthy of consideration. With regard to measurement strategy, this model adds a requirement to include measures of attributes whose dimensional meaning might be expected to change as a function of age.

Thus far I have been discussing a variety of developmental models arising from trait theory. An alternative conceptualization of development is found in the thinking of stage theorists, leading to a different approach to longitudinal data. The stage theorist believes that development consists of a series of qualitative changes in the organization of behavior. Longitudinal research is important here because it makes possible an empirical test of sequential orderings in stage progressions. Starting with a conception of each stage, measurement tries to detect the patterns of behavior characterizing each stage. Since these patterns are presumed to change with age rather than remaining invariant over time, the measurement strategy of the stage theorist differs from that of the trait theorist. For example, rather than seeking to maximize individual differences at each period of measurement, the stage theorist will attempt to maximize age changes in central tendency. In some instances these two strategies will even work at cross purposes! More critically, there is one variant of stage theory that would predict total lack of trait stability over time. Suppose that the processes facilitating or retarding stage progression differ at each stage and are uncorrelated. Under these conditions the stage theorist would predict no stability over time in stage-specific characteristics.

Differences among theories run even deeper, affecting measurement of environmental determinants as well as behavior. To illustrate this point, consider the contrast between a trait and stage conceptualization of environmental influence. A trait theorist might look for those environmental conditions and contingencies that mold the child's responses along certain channels rather than others. For example, he might look at reinforcement patterns in the home, or the availability of different types of imitation models in the home,

school, and peer group. Once these forces have acted upon the individual for some period of time, they presumably will determine certain traits that remain relatively fixed throughout life. In contrast, stage theory typically considers only one track rather than multiple dimensions. Each stage presumably is influenced by the environment, but environmental determinants do not ordinarily fixate individuals at a particular stage because progression rather than fixation is the rule. Here, the environment would appear to play an altogether different role. Instead of molding the individual to assume relatively fixed positions on multiple dimensions, the environment functions to accelerate or retard progressions along a series of qualitative reorganizations. Of course, patterns of reinforcement, social models, and other environmental influences could be quite relevant to this process, but their role is different. Whereas for the trait theorist any environmental impact is directly formative, in stage progression the environment functions either to support or suppress unidirectional developmental trends. This difference in conceptualization leads to different kinds of environmental measurement.

In this brief discussion I have presented only a few candidate models of human development, all calling for longitudinal research. To recapitulate the main argument, I started by saying that developmental theory can make a direct contribution to the conduct of educational research. My second point was that longitudinal designs offer an important and often crucial method for studying developmental phenomena. Finally, I have suggested that longitudinal designs provide a variety of potential virtues, none of which can be realized until specific models of development are carefully explored and linked to strategies of measurement.

Can You Do Real Research in the Real World?

Samuel Messick, Educational Testing Service

Is it possible to do real research in the real world? The answer is "Of course!" – but it's not easy. Not nearly as easy as doing real research in an artificial world, such as that provided by many laboratory settings. And even in the laboratory, where the application of various experimental controls makes specific interpretations more plausible, we sometimes pay a high price for this interpretability in the form of limited generality. An experimental treatment whose effects are evaluated under controlled conditions, for example, may not, because of reactions to the experimental conditions themselves, operate in the same manner in nonexperimental settings. The influence of work load on temper and interpersonal relations might turn out to be negligible during a simulated space flight in Houston, for instance, but not on board Apollo VII. Some results typical in the laboratory may thus not be typical in real life.

In choosing strategies for doing real research, then, whether in the real world or in the laboratory, we should ask not only how interpretable the results are likely to be but also how generalizable. Indeed, it is to variations in the degree of just these characteristics of interpretability and generalizability that we refer when we speak here of research as being more or less "real."

Generalizability and interpretability are two separate, though interrelated, issues. As we have already seen, laboratory findings may be clearly interpretable as due to the operation of a specific treatment, but the experimental conditions themselves may so color the responses as to severely limit generalizability to nonexperimental applications of the treatment. This particular threat to generalizability, incidentally, is pervasive and is not necessarily eliminated simply by avoiding the laboratory or controlled conditions. It may operate even in natural settings whenever the observer intrudes upon the scene, as in the celebrated "Hawthorne effect," and is one of the critical reasons for seeking, wherever possible, unobtrusive and nonreactive measurement conditions (1).

Because of this possibility of reactions to features of the experimental setting (and because of possible interactions between subject characteristics, such as intelligence or attitude, and conditions of the experiment), it becomes important in considering the applicability of the findings, and ultimately in interpreting their meaning, to ascertain in what other settings the effect will operate.

Similarly, subject characteristics may interact with the experimental treatment to produce different results for different kinds of people, so that it also becomes important to ask what other populations or types of subjects the results can be generalized to. This investigation of generalizability, whether across settings or populations or materials or whatever, is important not only to determine the range of applicability of the results but also to understand the nature of the results. Evidence for generality and for limitations in generality has a direct bearing on the interpretation of the findings, since it helps to specify those variables which, singly and in interaction, are necessary to produce the effect.

Another major question of generalizability – particularly in view of increasing recognition of the investigator's social responsibility to be alert to possible side effects in social science research – asks whether the effect is limited to particular measures in the intended outcome or whether it generalizes to other outcome measures: whether the adoption of a new mathematics curriculum in the early school years, for example, is associated not only with improved problem solving skills as intended but also, perhaps, with poorer computational skills, and perhaps not at all with changes in attitudes toward mathematics.

Another salient dimension of generalizability is the extent to which the effect can be generalized to other treatment variables – a question of special concern with complex treatments, such as curriculum programs or psychotherapies, as we attempt to determine what particular treatment variables or program components an effect may be attributed to.

In many instances in the real world, of course, we do not bother very much at all with evidence for generalizability – as, for example, when we wish to evaluate the effectiveness of the new third-grade remedial reading program in Franklin Elementary School during the spring term. Such a study is primarily concerned with describing the particular state of affairs for a given group of children receiving a specified treatment in a single setting during the chosen time period. Valuable as the study is for its delimited purpose of evaluating specific outcomes, we are offered little basis for deciding about the applicability of the treatment to other schools or to other types of students – although we may be willing to apply it anyway in the absence of other evidence – and we are at a loss to know how to modify the treatment if conditions change.

To meet these broader objectives, we need to undertake more comprehensive studies that compare observed effects across variations in setting, variations in type of subject, and variations in treatment

components. Furthermore, if these studies were also to include multiple measures of outcomes as well as multiple measures of subject characteristics, of background factors (including family, school, community, and peer-group influences), and of treatment components (including, in the case of educational programs, measures of teacher characteristics and classroom processes), we would then gain immense leverage on the problems of interpretation and generalizability – but this anticipates the argument somewhat. The point here is that as we expand our evaluation study from a description of effects for a particular group receiving a fixed treatment in a single setting to a study of differences in effect as a function of systematic variation, we add tremendous power to our research armamentarium. We are able to go beyond the particular case and generalize, to go beyond the specification of *what* is happening and infer *why* it happens – in short, to go beyond the descriptive to the scientific (2).

The key requirement in this enterprise is to be able to attribute observed effects to treatment components, whether directly or as interactions with other variables. In the simplest case, we need to be able to attribute an obtained effect – such as higher average reading scores at the end of a remedial reading curriculum than at the beginning – to the operation of the treatment under study and to rule out plausible rival hypotheses for explaining the gain, such as normal growth during that time interval, or practice effects from taking the pretest, or the occurrence of some other event (for example, a home reading program initiated by the school library during the same period). This is the basic problem of interpretability, and in the behavioral sciences it is usually resolved by using experimental designs employing control groups subjected to identical conditions except for the treatment.

In the logic of experimental design it is critical that treatments be assigned to subjects in complete independence of their prior states, so that the group of subjects receiving the experimental treatment does not differ initially in any systematic way from the control groups. This independence of treatment and prior state is effectively realized in practice by randomly assigning subjects to treatments. Under these conditions, if a significantly greater gain in reading scores is obtained in the treatment group than in the control group, the effect cannot be attributed to the occurrence of outside events or normal growth during the period, or to testing effects, or even to differential rates of maturation, for all of these should be comparable for the two groups.

In the real world, however, it is frequently difficult or impossible to use randomization procedures to establish comparison groups,

particularly in the study of certain ameliorative treatments that, for ethical or political reasons, cannot easily be withheld arbitrarily from the intended recipients. This tends to be the case with medicine and psychotherapy, for example, and with social betterment programs like Project Head Start. Although a strong moral case can be made for the use of randomization in the allocation of scarce resources that everybody needs, as was done in testing the Salk vaccine, such a rationale for access to limited social resources like compensatory education might not be nearly as acceptable politically as degree of need or timely enrollment. In addition, many practical reasons make it difficult to use randomization to study treatments in the context of social institutions — for example, the subjects, as part of a functioning system, are often already assigned to groups, like schools or classrooms, that are not easily disrupted (3). Furthermore, in some voluntary programs like Head Start, self-selection might lead to initial differences between those who attend and those eligible subjects who do not wish to attend on dimensions like desire to learn or parental encouragment, which might interact with treatment variables to produce greater gains for some than for others. In such a case, random assignment of *elibible* subjects to treatment and control groups might water down mean outcome differences and reduce generalizability to the natural setting. From this viewpoint, randomization would be desirable only within the applicant group, under circumstances where there are more applicants than openings.

Important as randomization is for experimental inference, its absence in a given study is no cause for despair. It is still possible to set up treatment and control groups that, although not strictly equivalent, will nonetheless be helpful in rendering many rival explanations implausible. The use of such nonequivalent control groups in the evaluation of treatment effects has been called a quasi-experiment, and the logic of quasi-experimental design, which has been discussed in detail elsewhere by Donald Campbell and others, provides a valuable rationale for much social science research (4).

Although these designs cannot be considered at length here, a few general principles may be summarized. One of the most popular quasi-experimental designs is the nonequivalent control-group procedure, which helps to attenuate all of the plausible rival hypotheses mentioned earlier except the possibility that greater gain in the treatment group might have been due to a different rate of maturation than in the control group. Initial differences between treatment and control groups due to selection biases, as well as differential attrition in the comparison groups, is handled statistically

by using gain scores or covariance methods. Thus, much of the design's value stems from the fact that *differences* in scores obtained by the *same* subjects at two points in time are compared across two groups, one receiving the treatment and one not. The power of this design is increased substantially if it is extended into a multiple time-series, having repeated measurements of the two groups over time with the treatment occurring for one group at some stage within the series. In this case a between-groups comparison of growth rates during nontreatment intervals with growth rates during the treatment interval provides a basis for evaluating the plausibility of differential maturation as a rival hypothesis for the treatment effect. This design can be naturally generalized to include more than two groups, whereby it becomes a longitudinal study of several groups exposed to different treatment alternatives.

One difficulty with quasi-experimental designs is that the effectiveness of control varies as a function of the similarity between the experimental and control groups in terms of both pretest scores and methods of selection. On one hand, we have seen how the experimentalist achieves effective control by using randomization to cut the causal strands of prior influence that might codetermine both exposure to the treatment and rate of change – as in the case of youngsters who attend Head Start classes because their parents want very much for them to learn. But is there any alternative or adjunct to randomization that would help us locate the critical tangled threads of interaction among prior influences and follow them as they become further enmeshed with other strings being pulled by treatment and background factors? The answer is "yes" – through the use of multiple measurement and multivariate analysis of covariation.

Multiple measurement is important even in true experimentation, for interactions due to unmeasured variables will not be properly taken into account with or without randomization. But it is particularly valuable when using nonequivalent control groups, for an attempt can then be made to specify the noncomparability in detail and to trace its possible consequences. With this general approach, we would endeavor to relate measures of subject variation and background variation to differential outcomes within treatment groups and to compare these relationships across groups as a function of measured treatment components. For example, in evaluating Head Start programs within this framework we would not only ask whether greater average gains on cognitive and personal-social dimensions are obtained for subjects exposed to the program as opposed to those who were not (and whether this effect holds for various subject

breakdowns, such as by sex or race or geographic region), but also what are the components of preschool education that are associated with growth in cognitive and personal-social functioning, and what are the individual and background factors that moderate these relationships.

To borrow a metaphor from Cronbach (5), the experimentalist is an expert puppeteer, able to keep untangled the strands to half-a-dozen independent variables. But in real life we are mere observers of a play in which Nature pulls a thousand strings and all the puppets are part Pinocchio. Multivariate analysis gives us a basis for figuring out where to look for the hidden strings – including those controlled by the puppets themselves – that animate the dance.

Notes

1. For a detailed discussion of this problem, see Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L. *Unobtrusive measures: nonreactive research in the social sciences.* Chicago: Rand McNally and Co., 1966.

2. For a further discussion of "specific evaluation" studies to describe what the effects are in contrast to "scientific evaluation" studies to infer why the effects occur, see Stake, R. E., Two approaches to evaluating instructional materials. (Paper delivered at the symposium on *Evaluation of Educational Materials and Processes,* American Psychological Association Meetings, San Francisco, August 30, 1968.)

3. For a discussion of policy positions that would make possible a greater use of randomization in field studies, see Campbell, D. T., Reforms as experiments. Evanston, Ill.: Northwestern University, unpublished manuscript, 1968.

4. Campbell, D. T. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin,* 1957, 54, 297-312; Campbell, D. T. From description to experimentation: interpreting trends as quasi-experiments. In C. W. Harris (Ed.), *Problems in measuring change.* Madison, Wis.: University of Wisconsin Press, 1963, 212-242. Campbell, D. T. Quasi-experimental design. In D. L. Sills (Ed.), *International encyclopedia of the social sciences.* New York: Macmillan Co., and Free Press, 1968, 5, 259-263; Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963, 171-246.

5. Cronbach, L. J. The two disciplines of scientific psychology. *American Psychologist,* 1957, 12, 671-684.

The ETS-OEO Longitudinal Study of Disadvantaged Children

Scarvia B. Anderson, Educational Testing Service

Educational Testing Service (ETS), under the auspices of the Office of Economic Opportunity (OEO), is embarking on a comprehensive study of the cognitive, personal, and social development of disadvantaged children over the crucial period from age 3 to grade 3. In very general terms, the aims of the study are to identify the components of early education that are associated with children's development, determine the environmental and background factors that influence such associations, and, if possible, describe *how* these influences operate. We hope to be able, eventually, to suggest what kinds of programs educational institutions might consider to bridge the gap between the disadvantaged and the more affluent, and to provide other information useful to community and federal planning agencies involved in problems of the poor.

Before we get into details of the plans for this ambitious study, however, let us take a look at what the target population is like. Actually, "target population" seems a very cold term for some 2,000 children who are about three and a half years old as the study gets under way.

Because of the particular concerns of the investigators and the sponsor, the children are poor. Many of them are black. Now you've heard all of the negatives about subjects like these: They live in city ghettos or rural shacks. They play with strings and boxes instead of the latest items from Creative Playthings. Sometimes one or both parents are missing from the home; frequently the parents are not what would be described in middle-class jargon as "satisfactory models." At best, they may project an image of defeat and helplessness. A few of the children may actually have brain damage; many of them suffer from malnutrition or lack of attention to correctable disorders. The language they speak and hear spoken is more than unacceptable – it is uninterpretable to many of us. And we throw up our hands in horror at the thought that a color TV set may rate higher on the family scale of values than proper food, clothing, or bedding.

But these children have two very powerful things going for them. First, they are eager, curious, and *young* – young enough that it's still possible to lay in them some kind of foundation for a good life. Second, most of them have some adult or adults in their lives who want more than anything else for things to be *better* for their

children. And they lend tremendous emotional – if not always intellectual – support to this aim.

Education is viewed as the major way to implement the aim. For the majority of children in the study, parents will make sure that they attend an educational program at the earliest possible opportunity. That educational program is known nationally as Head Start.

Mr. Messick has argued that, in spite of difficulties, it is possible – even essential – to do real research in the real world. However, the complexities of the design of the ETS study may cause you to wonder whether it's possible to do *it!*

It involves 9 groups of children in 23 elementary school sending districts in 4 geographical locations. The candidate locations are three cities varying in size, stability of the population, and degree of organization of the Negro community, and one rural-small town area in the South. All of the locations have Head Start available but the general outlines of the programs vary, reflecting the structural and curriculum differences of programs around the country. The nine groups of children in the study are listed in Table 1. (See page 32.)

To obtain the major subjects of the study – group 1 – we shall enter the designated school districts in the spring of 1969, knock on doors, and try to locate every child who will be eligible to enter the first grade in the fall of 1971. Of course, participation by these children in the study will be dependent on parental permission and cooperation. The cross-sectional comparison groups will be chosen from the same locations with the cooperation of local school and Head Start authorities.

Let me try to summarize some of the principal features of the study design:

First, the plan relies upon "natural" rather than "contrived" groups – parent decisions about sending or not sending children to Head Start or kindergarten will be made in the ordinary way.

Second, the study subjects will be Negro or white children from English-speaking backgrounds. For feasibility reasons, we did not wish to add the complications and numbers which the inclusion of Mexican-American, Puerto Rican, American Indian, and other special subgroups would entail. We hope that comparable studies of these children can be undertaken in the future.

Third, where possible, we have selected racially mixed school districts and we have made a point of including at least one district in each location where there is substantial variability in socioeconomic status. To the extent possible, we have tried to insure that race and SES are not completely confounded. (Race and SES are

of special interest as we study the effects of different classroom mixes on children of both races and of both lower and middle classes.)

Fourth, the cross-sectional comparison groups (groups 5-9) are viewed as an important design addition, principally as they provide a source of baseline data against which to interpret longitudinal results. Comparisons should be especially relevant in communities experiencing major social changes or upheavals during the course of the study and with respect to the cumulative effects of compensatory education.

Fifth, the purpose of reassessing comparison group 4 is to study the effects on children's development of the assessment procedures themselves. In addition, comparison group 3 (children moving into the classes) will permit us to gauge the cumulative effects of different amounts of assessment over the period of the study. It is possible (but we hope it doesn't happen) that the ETS measurements could exert a greater influence on the children than some of the compensatory educational experiences. In any case, we need to find out.

Now once we have the subjects of the study identified, what measures do we want to take on them – and why? With all due respect to Mr. Gordon's point about the inseparability of cognitive, physical, and personal-social growth, for convenience we are thinking in terms of several classes of measures that will be employed throughout the study. (We hope that structural analyses will throw important light on *how* these are interwoven.) These broad classes of measures are listed in Table 2: measures of the family; measures of the child's physical, perceptual, cognitive, and personal-social development; and measures of the classroom, teacher, school, and community. (See page 32.)

The choices of what measures to emphasize and use are, of course, based, on a number of considerations. Let me mention a few:

First, the questions toward which the study is directed require repeated measures of related phenomena over time. We may choose to measure exactly the same kind of thing over time – for example, breadth of vocabulary and goal directedness from age 3 through grade 3. Or we may measure characteristics that are thought to be precursors of later abilities of interest – visual and auditory perception at ages 3, 4, and 5 and reading ability at grades 1, 2, and 3.

Second, although the study will not overlook the usual demographic and static variables of home and classroom (things like family income, teacher's years of experience), we want to place

extraordinary emphasis on process variables (for example, teacher-child and parent-child interactions). These are the areas in which we think there will be payoff.

Third, the criterion measures of the study will encompass both the objectives that preschool and primary programs claim for themselves and aspects of development that society and social science theory hold as important in the broader area of human functioning.

Fourth, to the extent possible, we shall get multiple sources of information about a phenomenon – for example, from tests *and* from observations.

Fifth, for many of the measurements, we shall give preference to unobtrusive and nonreactive measures – for example, observations of children's behavior in natural settings.

Sixth, since descriptions of results should be handled at a level of discourse and conceptualization above the "item" level, every attempt will be made to develop and use psychologically and educationally meaningful scales. Of course, throughout we want to use measures that meet acceptable professional standards of reliability, validity, and so on.

In passing, I have made reference to parent permission and school cooperation. But in a study of this sort concern with parent, teacher, school, and community relations is of far more than passing significance. It is the key to whether the study ever gets started and, once started, gets done. In particular, many residents and teachers in poor or black areas are tired of the clipboarded researchers who cavalierly invade their lives, are suspicious of research completely planned and controlled by those outside the community and the culture, and are impatient with the lack of returns to the community.

We have to accept the notion that we *can* get past their reservations and conduct research in such areas – otherwise our study is dead – but we feel we have a special obligation to make the research as *relevant* as possible. Some of our steps in this direction include provisions for getting advice on measurement content and procedures from people in the study communities; having people on the central project staff who have lived or worked in similar communities; pretesting our procedures in similar communities (and with similar children, parents, and teachers); mounting an intensive public information program about the study in each area; "feeding back" relevant information to parents, school people, and others during the course of the study; and recruiting, training, and paying local personnel to carry out most of the operations required. Of course, we're not just being nice; we think such steps are essential to the validity of the study!

In trying to cover so much ground in such a short time, I'm afraid I have put several carts before several horses. Thus it may strike you as consistent, if a bit peculiar, for me to review now some of the questions that all of this talk about subjects, measures, and communities is about. Our general objective, as I have stated, is to try to find out about the componets of early education that are associated with the development of disadvantaged children. Furthermore, we feel that descriptions of effects should go beyond general or average trends. We want to know which particular program characteristics are best for which particular kinds of children. Moreover, to provide information that will contribute to educational and social planning, theories of child development, and techniques of assessing young children and their environments, we hope the study will be able to:

- find out how children's characteristics are related to home and community characteristics, and what characteristics distinguish the Head Start child from the eligible child who doesn't go to Head Start

- identify the characteristics of preschool and primary school programs in the study communities, and how these are supportive of one another or are in conflict

- determine not only the immediately apparent effects of compensatory preschool programs but also the permanence of any such effects through the primary grades

- relate teacher characteristics to teacher behavior

- obtain information about mobile versus nonmobile families

- describe changes in the interrelationships and structure of children's abilities and characteristics over time

- develop new means of assessing children and their environments.

This is a healthy order, and it takes a healthy staff to attempt to pull it off. The ETS "we" to whom I have referred frequently this afternoon includes a project direction consortium of Albert Beaton, Walter Emmerich, Samuel Messick, and me, assisted by Samuel Ball; Joseph Boyd, Program Coordinator; Virginia Shipman, Measurement

Coordinator; Samuel Barnett, Field Coordinator; and at least three dozen psychologists, educators, and statisticians who serve as task force leaders and members. The Steering Committee includes Silvan Tompkins and, not incidentally, some of the speakers this afternoon: Mr. Smith, Mr. Tumin, and Mr. Gordon.

Table 1: Subjects

GROUP 1

Major Ss of the study (eligible for first grade in 1971-72) who stay in the study districts. They are identified in spring 1969 and followed intensively through grade 3. N ≙ 2000 in 1969, 1000 in grade 3.

GROUP 2

Major Ss who move out of the study districts but are still assessed once a year. N ≙ 850 in grade 3.

GROUP 3

Classmates of major Ss – children who move into study districts after initial identification of group 1. N ≙ 550 in grade 1, 950 in grade 3.

GROUP 4

Cross-sectional comparison group (comparable school districts), assessed in Head Start and again in grade 3 in study of effects of assessment procedures themselves. N ≙ 450 in HS, 250 in grade 3.

GROUPS 5, 6, 7, 8, 9

Cross-sectional comparison groups (same school districts) assessed in 1969-70: HS, K, grade 1, grade 2, grade 3. (It is considered desirable to pick up additional cross-sectional comparison groups across the educational levels of the study in 1973-74 in order to assess program changes.)

Table 2: Measures

Family, status and process – To be obtained from interviews and observation of parent-child interaction for children in group 1 at the time of identification and annually throughout the study. Family interviews will also be carried out for children in group 2 who move away from the study locations. For reasons of economy, only family status information will be obtained on children in comparison groups 3-9.

Physical – To be obtained from medical examinations for children in group 1 at the time of identification and periodically throughout the study. Such medical information as available from preschool and school records will be obtained for children in the comparison groups.

Perceptual, cognitive – To be obtained through tests for children in group 1 at the time of identification and annually throughout the study, and for children in all other groups annually or as long as they are in the study. Teacher and parent ratings of cognitive development will also be obtained where appropriate.

Personal-social – To be obtained from observations in free-play situations once children are in preschool, from test-like situations where appropriate, and from ratings by testers and teachers for all groups. Parents will also be asked to make ratings of children in groups 1 and 2.

Classroom, program and climate – To be obtained from detailed observation of teachers and children in the classroom, from global ratings by observers, and from teacher descriptions for all preschool and school classes attended by children in groups 1, 3, 5-9. Limited data in this domain will be obtained for groups 2 and 4.

Teacher, background, attitudes, abilities, goals – To be obtained through questionnaires for all teachers every year they are involved with children in the study. For children who move away (group 2), every attempt will be made to involve their teachers in providing this information.

School, climate and structure – To be obtained from observations and from questionnaires completed by teachers and administrators. In addition, parents of children in groups 1 and 2 will be asked annually to give their attitudes toward the schools and classes their children are in.

Community – To be monitored by local observers throughout the course of the study. Parents will also be asked about their perceptions of the community and their access to its power structure and facilities.

The Scientific and Social Significance of the Longitudinal Study of Disadvantaged Children

*John W. McDavid, University of Miami**

I hope to remain brief in my remarks, and merely to comment on my perspective of the significance of the major longitudinal study of early educational experiences that Educational Testing Service has undertaken. My perspective is a dual one: On one hand, I wear the hat of the behavioral scientist seriously interested in new discovery and development related to the educative process. On the other hand, for the last year and a half I have worn the administrator's hat in a role of responsibility for evaluation of the massive social experiment known as Project Head Start.

I have always preferred to recognize Head Start as a social experiment. It is a set of manipulations and interventions being carried out on a grand scale with a wide array of socioculturally disadvantaged children and families. It is grounded in theory and accumulated knowledge about the human developmental process, the educative process, and social and community organization. Head Start's goals and objectives (in terms of betterment of the conditions of early physical, intellectual, personal, and social development of socioeconomically limited children) have been defined clearly from its very inception in the White House Conference on the Disadvantaged in 1964, and its establishment as a part of the Office of Economic Opportunity in 1965. However, it is not a conventional cut-and-dried social action program in the sense that Head Start has never selected one specific set of methods or procedures as the singly prescribed means of achieving these objectives. Quite intentionally, Head Start has chosen to offer only general directives and suggested alternatives as guidelines for developing local programs.

Because Head Start itself is a social experiment in early childhood education, it is a particularly appropriate vehicle for implementing the scientific ideas advanced in this symposium today. Head Start's value as a social experiment rests solidly on the quality of evaluative data gained as the experiment is carried out. Such data, in turn, will answer basic research questions about early educational experiences. Thus, evaluation and research are the same thing in this endeavor. For Educational Testing Service, the longitudinal study is primarily a piece of research; for Head Start and the Office of Economic Opportunity, it is an evaluation exercise. But the objectives of both parties will be served well by the study as it has been planned.

---

* Dr. McDavid was formerly Director, Research and Evaluation, Head Start.

In developing plans for evaluating Head Start, we have long recognized the need for a careful long-range study of the program's impact on children and their families. But we have felt that only recently has the time become appropriate for launching such a study. For two important reasons, it would not have been practical to initiate such a major effort at the very beginning of Head Start in 1965.

First, a span of time was needed to acquire program stability – to permit local groups and agencies planning and operating Head Start programs to assess and diagnose the pressing needs of the children and families they would serve, and to muster and mobilize the resources necessary to meet these needs. We now feel that this initial phase has passed, and that there is sufficient stability within Head Start programs around the country to justify the longitudinal study now planned.

Second, in 1965 educational research at the preschool level was seriously hampered by methodological inadequacies – we lacked sound methods for investigating such critical variables as personal, social, and motivational development of the child, or for analyzing specific elements of curriculum content. A "tooling up" period has been necessary to remedy this deficiency in research methodology. We have attempted to focus Head Start's research program along these lines, and, in fact, ETS has worked closely with Head Start for two years on these problems. Although we still recognize serious limitations of methodology, and critical lacks remain to be met, we feel that our level of methodological sophistication now warrants undertaking the longitudinal study.

So Head Start has negotiated a contract with ETS that will permit us to begin this six-year study, and we contemplate renewal of the contract each year to see the project through to its completion. It is hoped that other interested parties and agencies who would benefit from the results of the study may be induced to join in financial participation along the way, since the project will be an expensive one in terms of financial and intellectual resources.

Basically, then, as a major research effort, the design is a joint one. Head Start has identified the populations and manipulated the major *independent variables.* That is, Head Start has designated a set of populations of socioeconomically disadvantaged families and has offered a set of manipulations to intervene into the early developmental progression of young children. These manipulations include: diagnosis of medical deficits and provision of treatment for them; provision of stimulation and remediation for early intellectual and motivational deficits; provision of opportunity for improvement

of the disturbingly low self-regard and aspirations of children and parents; and provision of training and opportunity for improvement of the wage-earning capacity of families, their attitudes toward and participation in community affairs, and their perspective on matters related to the educational achievement of their children.

Having initiated that part of the research design concerned with manipulation of independent variables, Head Start has now asked ETS to execute the assessment of critical *dependent variables* that we expect will reflect the impact of Head Start's intervention. These include changes in intellectual capacity, academic achievement, motivation and goal-setting, self-regard, and attitudes toward community and society. Together we propose to digest and interpret this array of data, and from it all to satisfy both critical social and critical scientific needs.

There is a clear social need for sound data to plan the rapidly expanding range of massive federal involvement in early child development and services to children and families in several agencies of the Government. Furthermore, the basic scientific information gained here will facilitate our understanding of the general process of early child development. We will learn a great deal about the integration of intellectual, motivational, emotional, and interpersonal aspects of the child in his overall pattern of development. We will learn more about the characteristics of a hitherto little-recognized segment of our population who live their lives outside the mainstream of middle-class America, insulated from great segments of our culture. And we will learn more about the educational process itself, about its component elements, and why it works – or fails to work.

Mr. Smith raised a critical question in discussing the issue of continuity and discontinuity in early schooling, and I hope that the proposed longitudinal study may help to illuminate that question. There are currently several theoretical positions bearing on the importance of "early education," if we define education broadly to include all conditions designed to facilitate intellectual development. Some behavioral scientists hold that the preschool period represents a kind of "critical period" during which more or less irreversible damage to intellectual development may occur if there are deficiencies in environmental stimulation and opportunities to learn. Others, however, regard the developmental process as cumulative, with each succeeding stage building upon the prior. Before planning for effective education can even begin, we need greater information to determine whether Head Start should be construed as a one-shot effort to provide conditions otherwise lacking at a critical early period of development – or as merely one early step in a planned

and continuing effort to improve the educational environment for socioculturally disadvantaged children. Our early evidence from studies of Head Start so far strongly suggests the latter model.

There is a second way in which questions about the continuity of educational practice are critical. Some educators have traditionally held that "what is sauce for the goose is sauce for the gander." That is, "good educational practice" is regarded as good practice for everyone. This position generates efforts to find the recipe for the ideal curriculum, apart from any concern about those with whom it is to be used. An alternative position holds that good education is individualized – tailored to carefully diagnosed specific needs and capacities of the learner. In a sense, all good education is "special education." This position, then, generates efforts to relate specific curricular elements to specific learners. Head Start has generally been planned on the latter premise, but we certainly need additional sound data for further development of guidelines and directives.

Mr. Tumin's paper drew attention directly to an issue that has been at the heart of Head Start from its inception. Head Start has always strongly advocated expansion of the educational arena far beyond the boundaries of the classroom. The sociocultural context is recognized as a major determinant of early development, and Head Start has argued loudly to overcome traditionalism that circumscribes the role of education to the formal classroom. Head Start has attempted to work effectively with all facets of the child as a human being, and to intervene directly with his family, his neighborhood, and his community in order to provide improved circumstances for early development. This comprehensive concept of Head Start can be no more eloquently stated than in the words of Mr. Gordon (who has been identified from the beginning with planning Head Start's research and evaluation program) when he discussed the "false trichotomy" separating cognitive, motivational-emotional, and physical development of the child.

Mr. Emmerich's comments outline the ideal relationship between basic and applied research, or between good scientific investigation and useful program evaluation. Good program planning must be based on sound theory, and our only way of judging the soundness of theory is through careful empirical research or evaluation. Miss Anderson succinctly summarized a number of the most critical questions raised in the proposed longitudinal study for such careful empirical scrutiny. There is no doubt that we have a meeting of the minds at the level of scientific idealism in planning the longitudinal study!

But Mr. Messick's remarks have a sobering effect when he brings us

back to the work-day world by focusing attention upon obstacles that may make difficult the implementation of the ideal research design on which we all agree. Head Start is very much a "real world," and it is the arena in which we propose to conduct the beautiful research we have dreamed up. This investigation represents what Kurt Lewin called "action research," in that our experimental manipulations are producing very real consequences for very real people, and there is no insulating fence or boundary around this laboratory to slow the effects of these manipulations on all facets of their lives. We must be prepared for not only the expected, but the unexpected consequence as well. We must be prepared for making decisions that may represent compromise between the priorities of OEO with respect to program evaluation and those of ETS as a research organization. These priorities must be wedded in day-to-day decisions.

For example, we expect that Head Start may be seriously concerned with sample selection, since we should like to be able to vouch that the sample represents the full range of variation across the nation among Head Start programs and participants, so that we can generalize our findings and their implications. On the other hand, the nature of this study prohibits a large sample, and issues of feasibility, expediency, and cost may necessarily distort the representativeness of the sample. The urgent need for data on some issues or dimensions may require acceptance of methodological approaches that are too crude and subject to error to merit the most rigorous levels of scientific respectability. The fact that Head Start is embedded within a broader context of social action programs in the Office of Economic Opportunity may preclude opportunity for certain needed kinds of control and manipulation in order to frame research questions properly. The very fact that responsibility for Head Start clientele selection and program planning is ultimately lodged at the lowest administrative level (local programs) makes the coordination task extremely difficult, and magnifies problems of identifying proper control and comparison groups.

In summary, then, we do not expect execution of the longitudinal study to be without problems, and we are prepared to compromise when necessary to achieve the interests of both major parties involved. My perspective is such that I believe firmly that good basic research and good program evaluation can be integrated, and I am extremely pleased to have been a part of planning and developing this longitudinal project. But it is important that all of us recognize that Rome was not built in a day, and that no one single study, no matter how massive, can ideally provide all critical needs of both the

scientific community and the federal bureaucracy. From Head Start's point of view, the longitudinal study must be seen in perspective as but one of many large endeavors to evaluate Head Start as a federal social action program. In other studies we will focus on other facets of Head Start, we may have better opportunities for more comprehensive description of program or population variations, and we may have access to more representative samples. Program planning in Head Start will rest heavily – but not exclusively – on the results of this longitudinal study, and the administrative judgments there will, I trust, continue as they have in the past to reflect sound respect for good scientific evidence and efforts to integrate data from a wide variety of sources. In the same manner, I trust that all of us scientists recognize that although the ETS longitudinal study is potentially the most significant single piece of educational research undertaken in this decade, it must certainly be accompanied and followed by other equally ambitious efforts if we are eventually to meet our urgent social needs for sound educational theory and practice.